Description of Statistical Investigations across the levels from Level One to Level Four

Investigations at Level One

A start may be made with investigations by giving the students specific questions to answer. However, when they chose the questions themselves the experience becomes more meaningful and enriching. Hence you might encourage them to ask their own questions from an early stage.

Young students' questions are likely to focus on themselves and the activities that they are engaged in. Possible questions at Level 1 include:
How do you get to school?
What type of shoes do you wear?
What is your favourite colour? lolly? sandwich filling?
What sorts of shells are at the beach?

Deciding what kind of data to collect and how to collect the information can be difficult for young students. However the more opportunities they get to decide how to collect the data the more they will be able to make sense of the data.

Much of the data collected at Level 1 will be real objects themselves. For example, the students will collect shells from the beach, or will gather together a pencil from everyone in the class. Once the data (objects) is collected together they can be sorted into categories ready for display. It is important that the students are involved in deciding how to sort the objects. Sorting is an excellent way to encourage students to think about important features of data that lead to classifications that make sense.

Data at Level 1 will also be collected in the form of pictures. For example, the students can understand that pictures of cars are more appropriate than the cars themselves for collection. Once more it is all part of the appropriate *thinking* of this strand if the students are encouraged to make this decision for themselves.

Displays at Level 1 follow a simple progression from real object displays to pictographs (pictograms). First use the objects themselves to form a real objects display. Our shoes example could involve sorting the shoes into groups: shoes with laces, shoes with buckles and shoes with Velcro. Alternatively, the students can collect and sort pictures of vehicles, say, into a display.

The follow-up discussion will involve the students making statements about the number of objects in each of the categories. This might then be connected with Probability by way of statements such as shoes with laces are more common than shoes with buckles or there are more white cars than blue in the carpark.

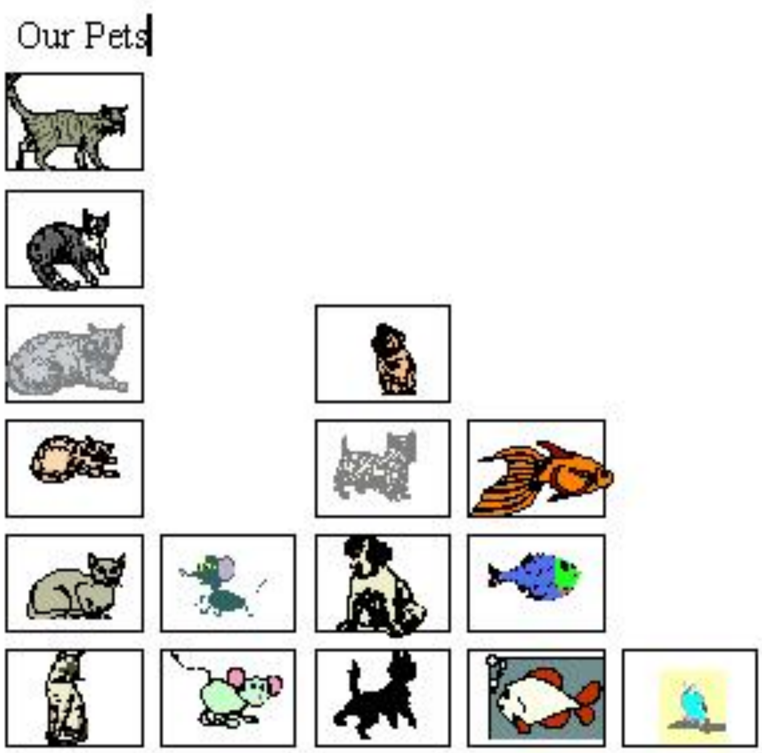
Investigations at Level Two

At Level 2 you can expect the students to be posing a greater range of questions. They will also be helped to understand some of the issues involved in conducting surveys and learn new methods for collecting data. While at Level 1 the students collected data and chose their own ways to display their findings, at Level 2 they will be introduced to uniform pictograms, tally charts and bar charts.

More emphasis here will also be placed on the discussion of the data and the making of sensible statements from both the student's own displays and the displays of others.

Uniform Pictograph

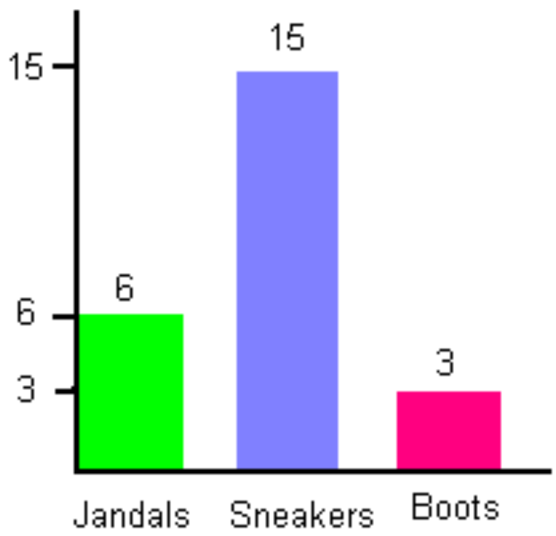
In a uniform pictograph the pictures are drawn on uniform pieces of paper. This means that the number of objects in each category now bears a direct relationship to the "size" of each category on the display. An example is shown in the diagram below.



In a further development the pictures can be displayed on a chart with axes and titles. The vertical axis can be numbered to match the pictures.

Bar Chart

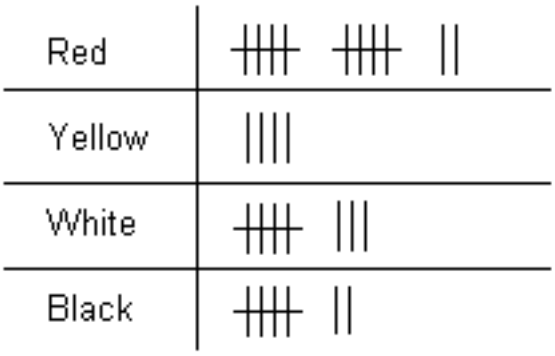
In a bar chart the pictures are replaced with vertical straight lines or rectangles. The position of these rectangles indicates what they represent and the height of these rectangles tells how many of that object there are.



The example above shows the types of shoes worn in the class on a particular day. There are three types of shoes: jandals, sneakers, and boots. The height of the corresponding rectangles shows that there are 6 lots of jandles, 15 lots of sneakers and 3 boots.

Tally Chart

A tally chart provides a quick method of recording data as events happen. So if the students are counting different coloured cars as they pass the school, a tally chart would be an appropriate means of recording the data. Note that it is usual to put down vertical strokes until there are four. Then the fifth stroke is drawn across the previous four. This process is continued until all the required data has been collected. The advantage of this method of tallying is that it enables the number of objects to be counted quickly and easily at the end.



In the example above, in the time that we were recording cars, there were 12 red cars, 4 yellow cars, 8 white cars and 7 black ones.

Investigations at Level Three

Planning at this level is more complicated than at Level 2. Here they can begin to talk about situations they have experienced, pose questions for an investigation and produce a plan for a statistical experiment. Children may be capable now of incorporating a computer into their work.

At Level 3 the students are introduced to stem-and-leaf graphs, dot plots and strip graphs.

Depending on the nature of the questions they are posing they will find out that they need to collect and organise the data in different ways. When considering the data they now start to see and talk about distinctive features of their displays such as outliers and modes.

Stem-and-leaf graph

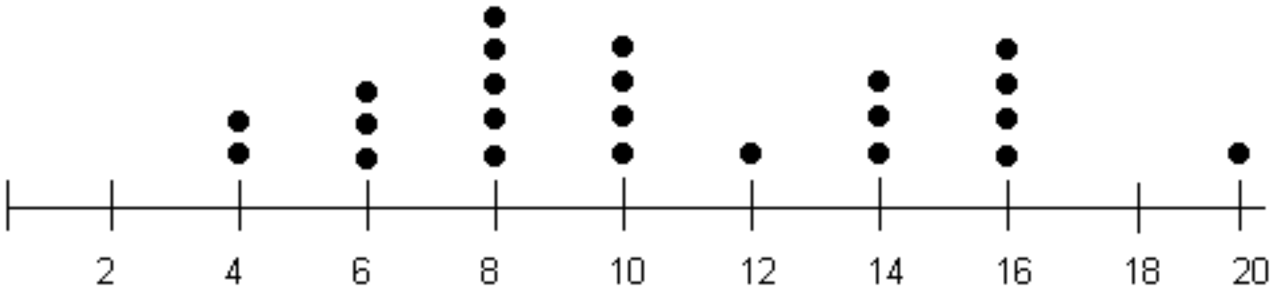
A stem-and-leaf graph is a bar graph made by arranging numerical data in a display. The first part of the number is the stem and the last part is the leaf. Data displayed in this way makes finding the range, mode and median (middle number) relatively easy.

Data: 12, 14, 17, 25, 36, 38, 40, 41, 43

1	2	4	7
2	5		
3	6	8	
4	0	1	3

Dot plots

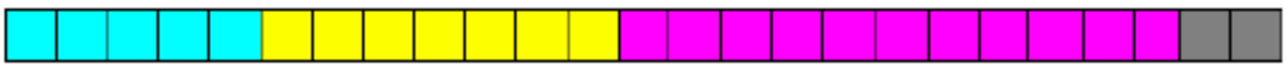
A dot plot represents data as dots on a scale. For example, in the dot plot below, we show the time students take to get to school from home. There are five students who take 8 minutes to get to school because there are five dots above the 8 minute mark.



Sometimes it is more convenient to show the dots as crosses. This is because dots can be more easily missed or deleted than crosses.

Strip Graph

A strip graph represents frequencies as a proportion of a rectangular strip. For example, the strip graph below shows that the students saw 5 light blue cars, 7 yellow cars, 11 maroon cars and 2 grey ones.



The strip graph can be readily developed from a bar graph. Instead of arranging the bars beside one another join them end to end. (Alternatively, you can easily get a bar graph from a strip graph by reversing the process.)

Pie Chart (from Strip Graph)

A pie graph is just like a piece of pizza pie divided up into sections. The size of the piece, that is the angle subtended at the centre of the pie, tells the relative magnitude of the object represented by that piece of pie. In the diagram, the light blue piece of pie represents 5/25 = 1/5 of the pie; the yellow section represents 7/25 of the pie; the maroon section, 11/25 of the pie; and the grey section 2/25 of the pie.

Actually this pie chart is easily formed from a strip graph. Take the strip in the example above and join the two ends to form a circle. Draw a circle this size on your paper. Then mark off the lengths of the different coloured parts on the edges of the circle. Join the ends of the marks to the centre of the circle. Of course reversing this process we can get a strip graph from a pie chart.

Investigations at Level Four

At the final stage, students should be motivated to plan their own investigation by some real world situation. They will then collect appropriate data and display it so as to emphasise its significant aspects.

Frequency tables

A frequency table shows the frequencies of data items in categories or intervals.

Eye colour	Frequency (number of students)
Blue	7
Green	4
Brown	8
Hazel	3

Back-to-back stem-and-leaf

A back to-back stem-and-leaf graph uses the same stem for two sets of data. The two data sets are then readily compared.

Pulse rate before exercise					Pulse rate after exercise				
8	3	2	0	5					
	7	6	2	6	6				
4	4	4	3	7	2	4			
	9	6	2	8	4	6	7	8	8
				9	2	4	5	9	
				10	0	2			

Mode, Mean and Median

Children at this level should be more sophisticated in their analysis and discussion of the data displays. They should be able to identify the mode, mean and median of sets of data. Discussions should lead to implications and actions consistent with the data.

The **mode** is the number that appears most often in a set of data. Suppose that Jenny scores the following points in successive nettle games: 4, 8, 4, 9, 4, 10, and 10. Then the most common of her scores is 4. This is the mode.

The number that all the numbers in a data set cluster equally around is the **mean**. This is calculated by adding all the numbers together and dividing by the number of numbers. In Jenny's example above the sum of all the numbers is 49. There are 7 numbers so the mean is 49/7 = 7.

The number that comes in the middle of a set of numbers when they are arranged in order is the **median**. In Jenny's case, if we order the data set we get 4, 4, 4, 8, 9, 10, and 10. The middle number here will be the fourth one. This number is 8, so the median of Jenny's scores is 8.

Now here we are lucky that there are an odd number of numbers. Otherwise there wouldn't be a precise middle number. If Jenny's scores had been 4, 4, 4, 8, 9, 10, then we have to take the three and a half number as the middle number. As this is halfway between 4 and 8 we take the mean of 4 and 8 to be the median. So the median in this case is 6.

The three concepts of mode, mean and median, measure central tendency in some way. That is, they give some idea of the "middle" number in a set. However, they are often different numbers. You can see this in Jenny's example above. The point of the mode is that its central tendency is the sameness of data, what is the most common same number. As for the mean, that tells which number is as close as possible to all numbers. When you add all the differences, both positive and negative, between the mean and the other numbers in the set, the result is zero. Finally the median is literally in the middle. When the data set is put in order, its the actual number that is at the halfway point of the list (or as close as we can get in the even case).

It is worth noting that all of these numbers can easily be found from stem-and-leaf plots. In the example above, every number occurs the same number of times so every number is the mode.

To find the mean, make a guess. Suppose that it is 25. Then the weight of the numbers above 25 is 11 (from the 36), 13 (from 38), 15 (from 40), 16 (41) and 18 (43) to give a total of 73. Below 25 we have 8 for (17), 11 (14) and 13 (12) to give a total of 32. The excess above 25 is 73 32 = 41. Since there are 9 numbers we divide this by 9 to get 41/9 = 4.56. So the mean is 4.56 above 25. In other words the mean is 29.6 to one decimal place.

The median is much easier. We just start at the bottom and count till we hit the 5th number (as there are 9 numbers here). This number is 36. So the median is 36.